COFIMA COMMAND REFERENCE

The program COFIMA (coordinate file manipulation) is a versatile program to make simple manipulations on Cartesian coordinate, distance constraint and dihedral angle constraint files. The program works interactively and allows for a variety of commands. Some of the operations that can be performed with COFIMA are:

- Convert between different data file formats
- Rename atoms, residues and dihedral angles
- Delete atoms, distance or angle constraints
- List specific atoms, distance or angle constraints
- Measure distances and dihedral angles
- Attach atoms (e.g. hydrogens)
- Insert pseudo atoms or pseudo atom constraints
- Generate covalent connectivities
- Sort atoms, distance or angle constraints

The program consists of three parts: COFIMA for coordinate file manipulations, DIFIMA for distance constraint file manipulations, and ANCOMA for angle constraints file manipulations. The prompts "`cofima>`", "`difima>`", and "`ancoma>`" indicate the part of the program that is currently active. Many commands can be used for all three types of data files, but there are also commands that are specific for certain types of data files. Commands can be abbreviated as long as the abbreviation remains unambiguous.

In the following description of the individual commands, A, A_1 etc. denote atom or angle names, R, R_1 etc. denote residue names, and r, r_1 etc. denote residue numbers. Atom, angle, and residue names must start with a letter and may (except in some cases) contain wildcards: "**`*`**" stands for

zero or more arbitrary characters, "**?**" stands for exactly one arbitrary character. No blanks are allowed within names. Residue numbers are integers. Atom, angle, residue names, and residue numbers may be preceded by an exclamation mark "**!**" which acts as a "not operator." The special residue names **FIRST** and **LAST** can be used to denote the first and the last residue in the coordinate file, respectively. The special residue names **first** and **last** can be used to denote the first and last residue of every fragment with contiguous residue numbers in the coordinate file, respectively. Atom, angle, and residue names (but not command words) are case-sensitive.

Many commands allow for the specification of a residue range, denoted by *range*, which consists of one or more of the elements "*r*", "*r***..**", "**..***r*", "r_1**..**r_2" or "**@***R*" (separated by at least one blank).

12	Residue 12
12 20..25	Residues 12, 20, 21, 22, 23, 24, 25
@THR	All residues with name "THR"
20..25 @THR	All residues with name "THR" and numbers 20–25
!@CY*	All residues with names that do not start with "CY"

The default residue range that will be used if no residue range is specified includes all residues.

For many commands all selected atoms must be in the same residue. This convention can be circumvented by preceding certain atom names with a tilde "**~**". In this case, atoms are searched through the list of covalent connectivities When using "**~**", covalent connectivities must of course be present; either they can be read from a DG coordinate file or they can be generated using the **connect**, **bind** or **link** commands.

The output of those commands that give interesting output can be redirected to disk files. To do this, the last parameter on the command line must be "**>**[*file*]" (here and in the following, items given in brackets are optional) which writes the output to a new file, or "**>>**[*file*]" which appends the output to an existing file. Note that no space is allowed between the **>** sign and the output file specification. If the output file specification is omitted, the previously used output file is used.

Sequences of commands that are often used may be stored in macros (different from INCLAN macros) with file name extension ".cfm" in order to facilitate routine applications of the program.

Macros can be called from within a macro. When executing a command, the program can detect two different types of problems, warnings which cause only the current command to be skipped, and errors which cause the whole rest of the macro to be skipped. Macros can be commented; text between the comment sign **#** and the end of a line is considered as a comment. A set of standard macros is provided with the program:

am_di Change from AMBER to DYANA nomenclature.

am_fm Change from AMBER to FANTOM nomenclature.

am_op Change from AMBER to OPAL nomenclature.

attach_am Attach hydrogens to amino acids and DNA, AMBER conventions.

backbone Keep only backbone atoms N, CA, C.

di_am Change from DYANA to AMBER nomenclature.

di_fm Change from DYANA to FANTOM nomenclature.

di_op Change from DYANA to OPAL nomenclature.

di_xp Change from DYANA to XPLOR nomenclature.

fm_am Change from FANTOM to AMBER nomenclature.

fm_di Change from FANTOM to DYANA nomenclature.

fm_pdb Rename residue names and last atom from FANTOM to PDB.

fm_xp Change from FANTOM to XPLOR nomenclature.

heavy Keep only heavy atoms.

norm_residues Achieve standard three letter code for amino acid residues starting from AMBER, DYANA, FANTOM or other reasonable residue names.

op_am Change from OPAL to AMBER nomenclature.

op_di Change from OPAL to DYANA nomenclature.

plimits Change upper limit distance constraints from real to pseudo atoms; DYANA nomenclature.

pseudo Insert pseudo atoms, DYANA nomenclature.

sort Sort atoms in amino acid residues.

If a command should only be applied to a certain type of data files, the command word may be followed (with no intervening spaces) by the qualifiers **/cofima** (to apply the command only to Cartesian coordinates), **/difima** (to apply the command only to distance constraints), **/ancoma** (to apply the command only to angle constraints), **/!cofima** (to not apply the command to Cartesian coordinates), **/!difima** (to not apply the command to distance constraints), or **/!ancoma** (to not apply the command to angle constraints).

The following, alphabetically ordered list of commands includes all commands that can be used for coordinate, distance constraint, and angle constraint files.

angles

[~]A_1 A_2 [~]A_3 [~]A_4 [[~]A_5] [*range*]

List bond angles, dihedral angles, or relative dihedral angles. This command can only be used with coordinate files. If three atom names are given, the bond angle A_1–A_2–A_3 is calculated. If four atom names are given, the dihedral angle A_1–A_2–A_3–A_4 is calculated. If five atom names are given, the difference between the dihedral angle A_1–A_2–A_3–A_4 and the dihedral angle A_1–A_2–A_3–A_5 is calculated.

For instance, the dihedral angles in a polypeptide can be calculated with the following command:

`angles CA C ~N ~CA` Calculate ω dihedral angles

ancoma

Switch to ANCOMA, the part of the program for the manipulation of angle constraint files.

attach

A [~]A_1 A_2 A_3 [[~]A_4] *b* τ θ [*range*]

Attach atoms to a structure. This command can only be used with coordinate files. The atom *A* is attached to the atom A_3 such that the bond length A_3–*A* equals *b*, the bond angle A_2–A_3–*A* equals τ, and the dihedral angle A_1–A_2–A_3–*A* (if A_4 is omitted) or the difference between the dihedral angles A_1–A_2–A_3–*A* and A_1–A_2–A_3–A_4 (if A_4 is present) equals θ (in this case it is not important which atom is specified by A_1). Note that *b*, τ and θ must be given as real numbers with a period to avoid confusion with the following *range* specification.

Normally, all atoms must be in the same residue. This convention can be circumvented by preceding atom names with a tilde "**~**". In this case, atoms are searched through the list of covalent connectivities which allows to use the **attach** command also if not all atoms lie within one residue.

`attach HB N CA CB OG1 1.09 110.9 123.0 @THR`

Attach the β-proton HB of threonine if the heavy atom positions are known.

bind

$A_1\ r_1\ A_2\ r_2$

Insert a specific covalent connectivity between the atom A_1 of residue r_1 and the atom A_2 of residue r_2. This command can only be used with coordinate files.

break

$A_1\ r_1\ A_2\ r_2$

Remove a specific covalent connectivity between the atom A_1 of residue r_1 and the atom A_2 of residue r_2. This command can only be used with coordinate files.

change

$A_1 \ldots A_2$ [*range*] @*R*|*r*|=*r*|+*r*|-*r*

Change residue names or residue numbers. If the last parameter is @*R*, the residue names of the specified atoms are set to *R*. If the last parameter is *r* or =*r*, the residue numbers of the specified atoms are set to *r*. If the last parameter is +*r* or -*r*, the residue numbers of the specified atoms are incremented or decremented by *r*.

cofima

Switch to COFIMA, the part of the program for the manipulation of Cartesian coordinate files.

connect

[A_1=$b_1 \ldots A_2$=b_2]

Generate covalent connectivities on the basis of bond length criteria. This command can only be used with coordinate files. Incorrect results may occur if there are large steric overlaps. The command without any parameters is equivalent to the following command:

```
connect H*=0.4 C*=0.85 N*=0.8 O*=0.7 S*=1.3
        P*=1.2 Q*=-999 LP*=-999 *=0.85
```

Usually, the command can be used with these default parameters. Covalent connectivities are generated for those atom pairs with the interatomic distance smaller than the sum of the two bond radii. The bond radius of an atom is given by b_i Å if A_i is the leftmost atom type on the command line that matches the atom name. Covalent connectivities are only generated between atoms that are in the same or in sequentially neighboring residues. To generate other connectivities, the commands **bind** and **link** can be used.

constraints $A_1 \ldots A_2$ [*range*]

List the distance or angle constraints involving the specified atoms or angles. This command can only be used for distance constraint or angle constraint files.

coordinates $A_1 \ldots A_2$ [*range*]

List the atom names, residue names and numbers, Cartesian coordinates, and, if present, covalent connectivities of the specified atoms. This command can only be used with coordinate files.

copy A_1 r_1 [A_2] r_2 [A_3]

Copy the atom A_1 of residue r_1, i. e. its Cartesian coordinates, to atom A_2 of residue r_2. This command can only be used with coordinate files. It adds a new atom to residue r_2. If A_2 is omitted, the name of the new atom will be A_1. If A_3 is given, the new atom will be inserted after A_3 in r_2, otherwise as the last atom of the residue.

delete $A_1 \ldots A_2$ [*range*]

Delete the specified atoms or constraints. When working with Cartesian coordinate files, all atoms whose name matches one of the atom specifications on the command line are deleted. When working with distance constraints, all distance constraints for which one or both atom names match an atom specification on the command line are deleted. When working with angle constraints, all constraints for angles whose name matches one of the angle specifications on the command line are deleted.

difima Switch to DIFIMA, the part of the program for the manipulation of distance constraint files.

directory

[*macro*]

Give a directory of all standard macro files and all macro files in the current working directory. If a *macro* specification is given, the directory will only contain those macro files with names that match the given *macro* specification. A *macro* specification is a macro file name, possibly containing wildcard characters, but excluding the extension ".cfm". For every macro, its name and the comment lines that precede the first command line are listed.

disconnect

$A_1 \ldots A_2$ [*range*]

Remove the covalent connectivities of the specified atoms. This command can only be used with coordinate files. The default is to remove all covalent connectivities.

distances

A_1 [$range_1$] A_2 [$range_2$] [*condition*]

When working with Cartesian coordinates, calculates the distances between atoms specified by A_1 $range_1$ and atoms specified by A_2 $range_2$. When working with distance constraints, list constraints for distances between atoms specified by A_1 $range_1$ and atoms specified by A_2 $range_2$. Optionally, only distances or constraints that fulfill one or several of the following *conditions* are listed:

d<*value*	distance less than *value*
d>*value*	distance greater than *value*
r<*value*	residue number difference less than *value*
r>*value*	residue number difference greater than *value*
r=*value*	residue number difference equal to *value*

Note that no spaces are allowed within a *condition*. For example, the command

```
distances HB% HN r=1 d<5
```

lists all sequential distances shorter than 5 Å between β and amide protons. The command cannot be used for angle constraints. end Terminate the program.

extract

A_1 [*range*$_1$] A_2 [*range*$_2$] [*condition*] [*limit*]

Extract constraints for the distances between atoms specified by A_1 *range*$_1$ and atoms specified by A_2 *range*$_2$. The extracted distance constraints are appended to the current list of distance constraints. Optionally, only constraints that fulfill one or several *conditions* are extracted. The format of a *condition* is the same as for the command **distances**. The distance limit is set according to an optional *limit* specification. It is possible to set the distance limit to the actual distance plus an offset by using the expression **l=***offset*, or to set the distance limit to the smallest possible value from a list of limits by using the expression **l<**l_1, l_2,..., l_n, or to set the distance limit to the largest possible value from a list of limits by using the expression **l>**l_1, l_2,..., l_n. Note that no spaces are allowed within these *limit* expressions. This command can only be used with coordinate files.

help

Display help information. Instead of help a question mark "**?**" may be used.

insert

A $A_1 \ldots A_2$ [*range*]

Insert pseudo atoms. This command can only be used with coordinate files. The command inserts a new atom with the name *A* in the centre of the atoms A_1,..., A_2. For example, the command

```
insert QB HB%
```

inserts a pseudo atom QB in the centre of the β-protons.

keep

$A_1 \ldots A_2$ [*range*]

Keep only those atoms, distance constraints, or angle constraints that match the specification given on the command line. As an example,

```
keep N CA C
```

deletes all atoms except the backbone atoms N, CA, and C' in amino acid residues. When working with distance constraints, distance constraints with one or both atom names matching an atom specification on the command line are kept.

link

[*b*] A_1 [$range_1$] A_2 [$range_2$]

Generates covalent connectivities between atoms called A_1 in the residue range $range_1$ and atoms called A_2 in the residue range $range_2$ if they are less than *b* Å apart. The default for bond length is b = 2.5 Å. This command can only be used with coordinate files. For example, the command

```
link SG SG
```

inserts covalent connectivities between atoms called SG which are less than 2.5 Å apart from each other, and can thus be used to generate the connectivities that correspond to disulfide bridges.

list

[*range*]

Gives a summary listing of the atoms, distance constraints or angle constraints in the given residue *range* (by default including all residues). The number of atoms, distance constraints or angle constraints, the number of residues, and lists of the occurring atom, angle, and residue names are given.

pseudo

A $A_1 \ldots A_2$ [A_3=c_3 . . .] ***=**c_4 [*range*]

Modifies distance constraints from real to pseudo atoms. This command can only be used with distance constraints files that contain upper distance bounds. Distance constraints involving atoms that match one of the atom specifications A_1,..., A_2 are changed in order to refer to the pseudo atom *A*, and the upper distance bound is increased by a correction. Usually, this correction is given by c_4 Å; if the distance constraint is an intraresidual constraint that involves one of the atoms A_3,... the specific correction given for this atom is used.

quit

Terminate the program.

read

file

Read an input *file* with Cartesian coordinates, distance constraints, or angle constraints. The program determines automatically which format the input data file has. The allowed formats for Cartesian coordinate files are:

DG	The format used by DYANA
PDB	The format used by the Protein Data Bank with some restrictions.
AMBER	The format used by the molecular dynamics program AMBER (very similar to PDB).

Distance constraints and angle constraints are read in the format used by DYANA.

remove

A_1 [$range_1$] A_2 [$range_2$] [*condition*]

Remove constraints for distances between atoms specified by A_1 $range_1$ and atoms specified by A_2 $range_2$. Optionally, only constraints that fulfill one or several *conditions* are removed. The format of a *condition* is the same as for the command **distances**. The command can only be used for distance constraints.

rename

A_1 A_2 [*range*]

Change the name of atoms or angles A_1 into A_2. As an example, the three commands

```
rename HB2 XXX
rename HB3 HB2
rename XXX HB3
```

exchange the names of the atoms HB2 and HB3.

retain

A_1 [$range_1$] A_2 [$range_2$] [*condition*]

Retain only constraints for distances between atoms specified A_1 $range_1$ and atoms specified by A_2 $range_2$. Optionally, only constraints that fulfill one or several *conditions* are retained. The format of a *condition* is the same as for the command **distances**. The command can only be used for distance constraints.

save

Write an output Cartesian coordinate, distance constraint, or angle constraint file with the same name and the same format as the input file from which the data was read.

sort

[$A_1 \ldots A_2$ * $A_3 \ldots A_4$] [*range*]

Sort atoms, distance constraints, or angle constraints. If there are no atom or angle specifications on the command line, the items are sorted according to a default order. Otherwise, the items are sorted into the order given by the atom or angle specifications on the command line. The asterisk "*****" represents all atoms or angles which are not explicitly given.

type

[*macro*]

List the contents of the macro file(s) that match the given *macro* specification. A *macro* specification is a macro file name, possibly containing wildcard characters, but excluding the extension ".cfm".

writeaco

file

Write an angle constraint output file in the format used by the program DYANA (see above). This command can only be used with angle constraint files.

writeamber

file

Write a Cartesian coordinate output file in AMBER format, the format used by the molecular dynamics program AMBER. This command can only be used with coordinate files.

writedco

file

Write a distance constraint output file in the format used by the program DYANA. The residue name and number of the first atom of the constraints are not repeated if they are the same as for the previous constraint. This command can only be used with distance constraint files.

writedg *file*

Write a Cartesian coordinate output file in DG format, the format used, for example, by the program DYANA. This command can only be used with coordinate files.

writelongdco *file*

Write a distance constraint output file in the format used by the program DYANA. The residue name and number of the first atom of all constraints are written out. This command can only be used with distance constraint files.

writepdb *file*

Write a Cartesian coordinate output file in PDB format, the format used by the Protein Data Bank. This command can only be used with coordinate files.

@macro Execute a *macro*, i. e. a file containing COFIMA commands. A *macro* specification is the file specification of the macro file excluding the extension ".cfm".

!*string* Repeat the last command that started with *string*. The *string* must not contain spaces.